

Pro-active medical information retrieval

Citation for published version (APA):

Braun, L. M. M. (2008). *Pro-active medical information retrieval*. Univeritaire Pers Maastricht. <https://doi.org/10.26481/dis.200810291b>

Document status and date:

Published: 01/01/2008

DOI:

[10.26481/dis.200810291b](https://doi.org/10.26481/dis.200810291b)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

Over the past years, various studies have proven that the retrieval of relevant, patient-related literature is vital to the quality of care. The number of medical errors has reached alarming heights and there is evidence that the use of information technology may reduce the number of medical errors. However, several obstacles in the medical domain obstruct successful medical information retrieval (IR).

In this thesis we investigate how to support physicians in the IR process. We believe that tailored IR support provides physicians with patient-related information and improves the quality of care. We investigate how to overcome two specific obstacles to medical IR: (a) the inadequate expression of information needs and (b) the time-consuming nature of the IR task. Consequently, we formulate a two-fold problem statement.

Problem statement (PS)

PS1: *To what extent can a physician's information needs – implicit and suppressed – be formulated automatically?*

PS2: *To what extent can the automatically formulated information needs be used as a starting point for the retrieval of relevant and patient-related literature?*

The problem statement results in five research questions. PS1 gives rise to three research questions (RQ1, RQ2, and RQ3), PS2 gives rise to two research questions (RQ4 and RQ5).

RQ1: *How can the information needs of a physician be modelled?*

RQ2: *How can patient-related information needs be formulated?*

RQ3: *How can different medical terminologies be mapped?*

RQ4: *How can a literature overload be prevented?*

RQ5: *How can the IR system be designed to be unobtrusive?*

The answers to the research questions are incorporated into a medical IR system: the Medical Information Retrieval Agent (MIRA).

In Chapter 2 we focus on RQ1: *How can the information needs of a physician be modelled?* We explain that modelling of information needs is an important step in the process of automatically formulating information needs. Without a model, it is impossible to handle all possible information needs that physicians may have. We use two distinct methods to identify the physicians' information needs, viz. a literature study and interviews with physicians. These methods result in a set of 181 identified information needs. Subsequently, the information needs are transformed into information-need templates: abstract representations of information needs. For the transformation, three methods are used: analysis, abstraction, and refinement. The application of these methods results in a set of 102 templates. These templates constitute INMOD, a model of physicians' information needs. INMOD is used to formulate a physician's information needs in a specific care situation.

Chapter 3 focuses on RQ2: *How can patient-related information needs be formulated?* and RQ3: *How can different medical terminologies be mapped?* In response to RQ2, we design an approach to bridge the gap from patient data to retrieved literature. The approach comprises five steps: (1) selecting templates, (2) extracting appropriate patient data, (3) translating patient data, (4) formulating information needs, and (5) retrieving literature. From the evaluation of MIRA we may conclude that our approach is feasible and can be generalized to other electronic medical records (EMRs), as long as they use a clear information structure that is programmatically accessible. However, the evaluation indicates that the number of formulated information needs per interaction is still high, as is the number of retrieved documents per interaction. In this way, physicians would be overloaded by information. So, we may conclude that our approach (up to this point) is not sufficiently adequate. In response to RQ3, we explain that Dutch non-standardized terms in the EMR present an obstacle to our approach. To address this obstacle, we investigate two different translation mechanisms, an automatic translation mechanism (AUTOTRANS) and a translation mechanism employing a manually constructed mapping (MANUTRANS). Experiments show that MANUTRANS provides us with the best translation results.

Chapter 4 focuses on RQ4: *How can a literature overload be prevented?* We investigate two approaches meant to prevent a literature overload. The first approach is to reduce the number of information needs. One way to achieve this is to discard information needs directly. There are four knowledge types that may be used to determine whether an information need should be discarded, viz. (a) formulation history, (b) entry order, (c) domain knowledge, and (d) number of retrieved documents. According to our experiments, the optimal formulation performance of MIRA is reached by using all four knowledge types as a discarding approach. The results of our tuning experiment indicate that this discarding approach reduces the average number of formulated information needs to a manageable number. The second way to reduce the number of information needs is to rank information needs according to their relevance and select the ten most-relevant ones (if more than ten information needs are formulated). There are three knowledge types that may be used to deter-

mine the relevance of an information need: (a) specialism, (b) entry time, and (c) number of retrieved documents. Optimal ranking performance is expected by using knowledge concerning the number of retrieved documents. The second approach to prevent a literature overload is to improve the precision of the retrieved documents. We investigate two precision enhancements: (a) patient-related document scoring and (b) patient-related document clipping. Optimal retrieval performance is expected when using patient-related document clipping as our precision-enhancement approach.

Chapter 5 provides an evaluation of MIRA and answers RQ5: *How can the IR system be designed to be unobtrusive?* The assessment of MIRA's performance is provided by nine physicians with three distinct levels of medical experience, viz. attending physicians, resident physicians, and recently graduated physicians. MIRA is evaluated with respect to four evaluation criteria: (a) retrieval performance, (b) formulation performance, (c) ranking performance, and (d) unobtrusiveness. With respect to the retrieval performance, we observe that the investigated precision enhancements (patient-related document scoring and patient-related document clipping) do not enhance the precision of the retrieved document set. Consequently, the final version of MIRA does not incorporate any of the precision-enhancement approaches. The evaluation of the formulation performance indicates that MIRA's performance is adequate according to the attending physicians and the resident physicians. For the recently graduated physicians, the results indicate that MIRA is capable of generating implicit information needs and retrieving documents answering the implicit information needs. With respect to the ranking performance, we observe that the ranking method to be used depends on the type of physician using MIRA. Attending physicians prefer ranking according to their specialism, resident physicians prefer ranking according to entry time of the patient data, and recently graduated physicians have no explicit preference for any of the ranking criteria. With respect to the unobtrusiveness of MIRA (in response to RQ5), we may state that MIRA can be designed to be unobtrusive by (a) preventing workflow interruptions, (b) preventing literature overload, and (c) preventing retrieval of non-relevant literature. The current version of MIRA is unobtrusive with respect to prevention of workflow interruptions. However, MIRA is obtrusive with respect to literature overload and the retrieval of non-relevant literature. Consequently, MIRA has to be improved with respect to the latter two aspects.

In Chapter 6 we use the answers to the five research questions to give an answer to our problem statement. With respect to PS1, we may conclude that it is possible to formulate a physician's information needs automatically. This statement holds for explicit, implicit, and suppressed information needs. With respect to PS2, we may conclude that the information needs formulated by MIRA can be used as a starting point for information retrieval under certain conditions. Finally, Chapter 6 discusses four topics with respect to our research and provides three directions for future research.

Samenvatting

De laatste jaren hebben verschillende studies laten zien dat de retrieval van relevante, patiënt-gerelateerde literatuur essentieel is voor de kwaliteit van de medische zorg. Het aantal medische fouten heeft een alarmerende hoogte bereikt en er zijn duidelijke aanwijzingen dat het gebruik van informatietechnologie dit aantal kan verminderen. Succesvolle medische *information retrieval* (IR) wordt echter gehinderd door verschillende obstakels in het medische domein.

In dit proefschrift beschrijven we hoe artsen ondersteund kunnen worden in het IR-proces. Wij denken dat gespecialiseerde IR-ondersteuning artsen kan voorzien van patiënt-gerelateerde informatie die kan leiden tot een verbetering van de kwaliteit van de zorg. In dit kader onderzoeken wij hoe twee specifieke obstakels van medische IR overwonnen kunnen worden: (a) de inadequate expressie van informatiebehoeften en (b) de tijdrovende aard van het IR-proces. Derhalve formuleren we een tweeledige probleemstelling.

Probleemstelling (PS)

PS1: *In welke mate kunnen de informatiebehoeften – impliciet en onderdrukt – van een arts automatisch geformuleerd worden?*

PS2: *In welke mate kunnen de automatische geformuleerde informatiebehoeften gebruikt worden als een uitgangspunt voor de retrieval van relevante en patiënt-gerelateerde literatuur?*

De probleemstelling leidt tot vijf onderzoeksvragen. PS1 leidt tot drie onderzoeksvragen (RQ1¹, RQ2, en RQ3), PS2 tot twee onderzoeksvragen (RQ4 en RQ5).

RQ1: *Hoe kunnen de informatiebehoeften van een arts gemodelleerd worden?*

RQ2: *Hoe kunnen patiënt-gerelateerde informatiebehoeften geformuleerd worden?*

RQ3: *Hoe kunnen verschillende medische terminologieën gematcht worden?*

¹De afkorting *RQ* is afgeleid van het engelstalige *Research Question*.

RQ4: *Hoe kan een overdaad aan literatuur voorkomen worden?*

RQ5: *Hoe kan het IR-systeem zo ontworpen worden dat het niet hinderlijk is?*

De antwoorden op de onderzoeksvragen worden opgenomen in een medisch IR-systeem: de Medische Information Retrieval Agent (MIRA).

In Hoofdstuk 2 richten we ons op RQ1: *Hoe kunnen de informatiebehoeften van een arts gemodelleerd worden?* We leggen uit dat het modelleren van informatiebehoeften een belangrijke stap is in het proces van de automatische formulering van informatiebehoeften. Zonder een model is het onmogelijk om alle potentiële informatiebehoeften van artsen te formuleren. We gebruiken twee verschillende methoden om de informatiebehoeften van artsen te identificeren, namelijk een literatuurstudie en interviews met artsen. Deze methoden resulteren in een totaal van 181 geïdentificeerde informatiebehoeften. Vervolgens worden de informatiebehoeften getransformeerd in informatiebehoefte-*templates*: abstracte representaties van informatiebehoeften. Voor de transformatie gebruiken we drie methoden: analyse, abstractie, en verfijning. De toepassing van deze methoden resulteert in een verzameling van 102 templates. Deze templates vormen INMOD, een model voor de informatiebehoeften van artsen, dat gebruikt wordt om genoemde informatiebehoeften te formuleren in een specifieke zorgsituatie.

Hoofdstuk 3 richt zich op RQ2: *Hoe kunnen patiënt-gerelateerde informatiebehoeften geformuleerd worden?* en RQ3: *Hoe kunnen verschillende medische terminologieën gematcht worden?* Met betrekking tot RQ2 ontwerpen we een benadering die een brug slaat tussen patiëntdata en literatuur. Deze benadering omvat vijf stappen: (1) selectie van *templates*, (2) extractie van de juiste patiëntdata, (3) vertaling van patiëntdata, (4) formulering van informatiebehoeften, en (5) retrieval van literatuur. Uit de evaluatie van de benadering mogen we concluderen dat deze haalbaar is en gegeneraliseerd kan worden naar andere elektronische medische dossiers (EMD, in het Engels: EMR), zolang deze een duidelijke informatiestructuur gebruiken die toegankelijk is voor MIRA. Daarnaast maakt de evaluatie duidelijk dat het aantal informatiebehoeften dat per interactie geformuleerd wordt nog erg hoog is, evenals het aantal documenten dat gepresenteerd wordt. Hierdoor zouden artsen een overdaad aan literatuur te verwerken krijgen. We mogen dus concluderen dat de benadering (tot zo ver) nog niet adequaat is. Met betrekking tot RQ3 laten we zien dat de nederlandsstalige, ongestandaardiseerde termen in het EMR een obstakel vormen voor onze benadering. Om dit obstakel te elimineren onderzoeken we twee vertalingsmechanismen: een automatisch vertalingsmechanisme (AUTOTRANS) en een vertalingsmechanisme uitgaande van een handmatig geconstrueerde projectie (MANUTRANS). Experimenten wijzen uit dat MANUTRANS tot de beste resultaten leidt.

Hoofdstuk 4 richt zich op RQ4: *Hoe kan een overdaad aan literatuur voorkomen worden?* We onderzoeken twee benaderingen die moeten voorkomen dat een arts een overdaad aan literatuur te verwerken krijgt. De eerste benadering is het verminderen van het aantal geformuleerde informatiebehoeften. Eén manier om dit te bereiken is het direct verwerpen van de informatiebehoeften. Er zijn vier kennistypes die

gebruikt kunnen worden om te bepalen of een informatiebehoefte verworpen moet worden: (a) formuleringsgeschiedenis, (b) invoervolgorde, (c) domeinkennis, en (d) aantal gevonden documenten. Uit onze experimenten blijkt dat MIRA een optimale formulering bereikt wanneer alle vier de kennistypes in de verwerpsmethode aanwezig zijn. De resultaten van deze experimenten geven aan dat het aantal geformuleerde informatiebehoeften op die manier wordt verminderd tot een aanvaardbaar aantal. De tweede manier om het aantal informatiebehoeften te verminderen is door ze te ordenen op grond van hun relevantie en de beste tien te selecteren (aangenomen dat er meer dan tien informatiebehoeften geformuleerd zijn). Drie kennistypes kunnen gebruikt worden om de relevantie van een informatiebehoefte te bepalen: (a) specialisme, (b) invoertijdstip, en (c) aantal gevonden documenten. Op grond van onze experimenten mogen we een optimale ordening verwachten bij gebruik van alleen het laatste kennistype. De tweede manier om te voorkomen dat artsen een overdaad aan literatuur te verwerken krijgen is het verbeteren van de precisie van de gevonden documenten. We onderzoeken twee precisie-verbeteringen: (a) de patiënt-gerelateerde document-score en (b) de patiënt-gerelateerde document-vermindering. We verwachten een optimale precisie bij gebruik van patiënt-gerelateerde document-vermindering.

Hoofdstuk 5 geeft een evaluatie van MIRA en beantwoordt RQ5: *Hoe kan het IR-systeem zo ontworpen worden dat het niet hinderlijk is?* De beoordeling van MIRA's prestatie wordt gegeven door negen artsen met verschillende niveaus van medische ervaring, namelijk specialisten, arts-assistenten, en basisartsen. MIRA wordt geëvalueerd op grond van vier evaluatiecriteria: (a) *retrieval*, (b) formulering, (c) ordening, en (d) hinderlijkheid. Met betrekking tot de *retrieval* merken we op dat de precisie-verbeteringen (patiënt-gerelateerde document score en patiënt-gerelateerde document vermindering) de precisie van de gevonden documenten niet daadwerkelijk verbeteren. Om die reden is geen van beide benaderingen opgenomen in de uiteindelijke versie van MIRA. De evaluatie van de formuleringsprestatie maakt duidelijk dat MIRA's prestatie adequaat is volgens de specialisten en de arts-assistenten. Met betrekking tot de basisartsen maken de evaluatieresultaten duidelijk dat MIRA in staat is om impliciete informatiebehoeften te formuleren en documenten te vinden die aan de informatiebehoeften beantwoorden. Met betrekking tot de ordening merken we op dat de te gebruiken ordeningsmethode afhangt van de arts (type arts) door wie MIRA gebruikt wordt. Specialisten geven de voorkeur aan een ordening op grond van hun specialisme, arts-assistenten geven de voorkeur aan een ordening op grond van het invoertijdstip van de patiëntdata, en basisartsen hebben geen expliciete voorkeur. Met betrekking tot de hinderlijkheid van MIRA (als antwoord op RQ5) kunnen we vermelden dat MIRA als niet-hinderlijk kan worden ontworpen door het voorkomen van (a) werkonderbrekingen, (b) een overdaad aan literatuur, en (c) het presenteren van niet-relevante literatuur. De huidige versie van MIRA is niet-hinderlijk met betrekking tot de werkonderbrekingen. Met betrekking tot een overdaad aan literatuur en de presentatie van niet-relevante literatuur is MIRA volgens sommige artsen echter wel hinderlijk. Hieruit blijkt dat MIRA nog verbeterd moet worden op deze laatste twee gebieden.

In Hoofdstuk 6 gebruiken we de antwoorden op de vijf onderzoeksvragen om een

antwoord te geven op onze probleemstelling. Met betrekking tot PS1 mogen we concluderen dat informatiebehoeften van een arts automatisch geformuleerd kunnen worden. Dit is het geval voor expliciete, impliciete, en onderdrukte informatiebehoeften. Met betrekking tot PS2 mogen we concluderen dat de informatiebehoeften die MIRA formuleert, gebruikt kunnen worden als een uitgangspunt voor IR onder bepaalde voorwaarden. Tenslotte bespreekt Hoofdstuk 6 vier onderwerpen met betrekking tot ons onderzoek en wijst het op drie richtingen voor toekomstig onderzoek.